Glossary of terms for Ensembles

Forecasting Research Technical Report No. 663

November 2024

Edited: Ken Mylne

Contributors: Oak Wells, David Walters, Adam Scaife, Nigel Roberts[1], Rosa Barciela

DOI: https://doi.org/10.62998/gwwk6803

1. European Centre for Medium-Range Weather Forecasts, Reading.

© Crown Copyright 2024, Met Office

Met Office

Document history

Version	Purpose	Date
0.1	Draft	31 July 2024
0.2	Revision	13 August 2024
1.0	Revision for external submission	21 August 2024
1.1	Revision after further review	24 October 2024
1.2	Final Tech Note format	19 November 2024

Prepared by

Ken Mylne, Oak Wells, David Walters, Adam Scaife, Nigel Roberts, Rosa Barciela

Reviewed by

Rosa Barciela, Head of Ensembles Exploitation, 05/08/2024

Keith Williams, Robert Neal, Diana Greenslade*, Harald Richter*

(* Australian Bureau of Meteorology)

Authorised for issue by

Rosa Barciela, Head of Ensembles Exploitation, 21/08/2024

© Crown copyright 2024, Met Office

Met Office

Glossary of terms for Ensembles

Purpose

To enable us to be clear and specific in our use of language around ensembles in our communications and across timescales. Work on this collaboratively with international colleagues to set an international standard for language around ensembles.

Use

Initially for use in strategic action on Exploiting Ensembles and in R&I theme on Producing and Exploiting Ensembles. To be expanded over time for use across Met Office and externally. Consequently, it is easiest if we are consistent with any accepted international usages of language within the Met community.

Glossary

Qualitative terms with proposed definitions

Analysis Perturbations: see Initial Condition Perturbations

Best/worst case: see worst case

Bi-modal (or multi-modal): empirical ensemble distributions can sometimes show probability distributions with two (or more) modes (peaks) due to diverging clusters of solutions.

Central member: the single prediction realisation which best represents the "centroid" of the full ensemble**;** the prediction time period over which the member is defined to be central must be clearly stated**.** See also **medoid**.

Clusters: Groupings of similar members within an ensemble. Clusters could be groupings of members that best associate with pre-learned possible outcomes (flow regimes or weather types), or members could be grouped according to a measure of how close they are. The prediction time period over which the clusters are defined must be clearly stated**.**

Control: A member of the ensemble which does not have any perturbations applied.

Covariance information/ Meteorological Covariance: individual ensemble members retain the meteorologically consistent covariance information between different weather elements (variables) and also spatio-temporal covariance information important for some applications. Such covariance information can be lost in some (calibrated) probability prediction representations.

Ensemble: A collection of prediction realisations for the same period or event.

Ensemble member: A single prediction realisation.

Ensemble of opportunity: see Opportunistic ensemble

Ensemble Prediction System (EPS): A weather and/or climate prediction system (or similar e.g. Ocean Prediction System) exploiting ensemble methods.

Ensemble subset: a smaller number of ensemble members with a common characteristic. This could be in the application of ensemble output to a product or could be a method of selecting which members should be used to drive some other system (e.g. a large global ensemble driving a smaller regional ensemble) or to examine conditional hazards or impacts.

Hindcast: A set of retrospective predictions (see also reforecast for more detail)

Initial Condition Perturbations (or Analysis Perturbations): Small changes made to the model initial state (analysis) representing analysis uncertainty in order to stimulate growth of ensemble spread.

© Crown copyright 2024, Met Office

Various methods may be applied to create perturbations which actively stimulate growing modes including use of singular vectors, forms of ensemble Kalman filter or time-lagged analyses.

Initial condition uncertainty (or analysis uncertainty): uncertainty in the prediction which derives from uncertainty in the initial state (analysis)

Medoid: The value of the member representing the central prediction in a collection of forecasts/members. If an ensemble is clustered, there can be a medoid for each cluster as well as for the full ensemble. Unlike the median or mean, the medoid value must come from one of the forecasts/members.

Model physics: components of a numerical model representing scale processes which are not fully resolved and typically simulated using parameterization schemes. Such processes can involve significant sub-grid scale uncertainty with the potential to lead to errors later in the prediction – see model uncertainty and stochastic physics. Is usually considered distinct from the 'dynamics' (representation of resolved fluid dynamics).

Model resolution: the degree to which a numerical model is able to explicitly resolve physical processes and which is largely determined by the grid-spacing of the model, both horizontal and vertical. In practice the term is frequently used to explicitly describe the (horizontal) grid-length. In reality the true resolution of a model has a length scale of several grid-lengths since several grid-points are required to simulate a physical process.

Model uncertainty/model error: uncertainty in the prediction which derives from uncertainty in the model representation of the atmosphere, primarily in the physics parameterisations. Model uncertainty complements initial condition uncertainty to sum up to the total forecast uncertainty represented by an ensemble forecast.

Multi-model ensemble: the combination of information/ensemble members from more than one ensemble prediction system. This has greater diversity, driven by multiple initial analyses and perturbation schemes, multiple dynamical cores and multiple physics packages and is used to generate a larger, more complete or more skilful ensemble.

Multi-physics ensemble: a method for representing uncertainty within physical parametrisation schemes in which different ensemble members use either different parametrisation schemes or significantly different settings within their parametrisation schemes.

Opportunistic ensemble/ensemble of opportunity: an ensemble formed by combining the output of several deterministic model predictions. These could use output from multiple centres (e.g. Met Office GM, ECMWF HRES IFS and NOAA GFS) or could be a deliberately formulated ensemble (e.g. NOAA's High-Resolution Ensemble Forecast system). Such systems have previously been described as a 'poor man's ensemble'.

Perturbations: Small differences introduced to ensemble members to represent uncertainty in the initial or model state. See also Initial Condition Perturbations and Stochastic Physics.

Perturbed Parameter Ensemble: a method for representing uncertainty within physical parametrisation schemes in which different members use different (but reasonable) settings within a fixed set of parametrisation schemes. Often used in climate change ensembles. This differs from a "random parameters" approach by using fixed (in space and time) parameter values within any given ensemble member.

Poor man's ensemble: see Opportunistic ensemble. (Use of this term is no longer encouraged.)

Predictable limit: the maximum forecast lead-time at which forecast skill is achievable given the limits of chaos theory.

© Crown copyright 2024, Met Office

Probability forecast: a forecast that specifies the likelihood of occurrence of a mutually exclusive and collectively exhaustive set of events during a given time frame. Probability forecasts range from 0.0 (event cannot occur) to 1.0 (event is certain to occur). May also be expressed as a percentage. (Taken from NOAA glossary.) An ensemble forecast is not a probability forecast but may be used to estimate probabilities.

Provenance of ensemble/member: Information/metadata used to describe the links between an ensemble (including its individual members) with the upstream systems used in its creation. For example, regional ensemble members driven by specific global ensemble members, or surge and wave predictions with a specific parent atmospheric model scenario.

Random parameters: see Stochastic physics

Realisation: alternative term for an ensemble member, representing a complete meteorologically consistent prediction evolution

Reasonable worst case: see worst case

Reforecast: A set of retrospective predictions designed to create a multi-year (or multi-decadal) model climatology with a current version of the model. Reforecasts are primarily used to assess where forecasts with the same current model are situated within their own model (or ensemble) climatology. This allows the classification of forecasts in terms of model-climatology anomalies, as a proxy for likely impact. (Reforecasts may also be referred to as hindcasts, particularly for longer-range forecasts.)

Representative member: The member which best represents the most likely outcome from the full ensemble. If an ensemble is clustered, there can be a representative member for each cluster as well as for the full ensemble.

Scenario: A single plausible forecast pattern predicted by one or more ensemble members (or artificially constructed from these) and distinct from separate patterns forecast by other members of the same ensemble. In practice a scenario may be represented by a single ensemble member.

Stochastic Kinetic Energy Backscatter (SKEB): see Stochastic physics

Stochastic Perturbations of Physical Tendencies (SPPT): see Stochastic physics

Stochastic Perturbed Parameters (SPP): see Stochastic physics

Stochastic physics: Methods for representing uncertainty or unmodelled variability within physical parametrisation schemes, sometimes termed “model uncertainty”. These currently include Random Parameters (RP, where global parameter values are perturbed in a time-varying way), Stochastic Perturbed Parameters (SPP, where parameters are perturbed in a spatially- and time-varying way), Stochastic Kinetic Energy Backscatter (SKEB) and Stochastic Perturbations of Physical Tendencies (SPPT) or Stochastic Perturbed Tendencies (SPT), where the tendencies diagnosed by the parametrisation are perturbed before being applied to the model. (See also Perturbed Parameter.)

Storyline: a possible evolution taken from an ensemble member or subset of ensemble members that describes a plausible course of events. Often involves an interpretation of a forecast by an expert meteorologist, sometimes for a lay audience.

Time-lagging: the combination of ensemble members initiated at different analysis times to increase ensemble size and/or improve performance. Time-lagging can also be used as a method of generating initial condition perturbations, particularly for long-range predictions.

Worst (best) case: an ensemble member at the extreme of the distribution which represents highest (lowest) impact for a particular user or application and may be used for contingency planning.

© Crown copyright 2024, Met Office

Sometimes termed Reasonable Worst Case recognising that reality could be an outlier worse than any ensemble member. May also be defined as an extreme percentile (e.g. 5th or 95th) rather than as an ensemble member.

Quantitative terms with proposed definitions

Analysis: an initial state for a prediction representing the best-estimate of the current state of the atmosphere/ocean, typically generated from a combination of recent observations with a background forecast through a data assimilation system.

Base rate: The rate of occurrence of the observations.

Bias/Ensemble Mean Bias: The difference between observations and ensemble mean for a given lead time, averaged over many cases. The degree of correspondence between the mean prediction and the mean observation. This type of bias is also known as overall bias, systematic bias, or unconditional bias. The mean error is a measure of the overall prediction bias for continuous and probabilistic forecasts (taken from NOAA glossary).

Calibration: the use of statistical or machine learning methods to improve the statistical performance of (ensemble) predictions based on past performance comparing predictions with observations. Effective calibration can generate valid probability predictions from ensemble estimates.

Centroid: The mean value of a collection of forecasts/members. If an ensemble is clustered, there can be a centroid for each cluster as well as for the full ensemble.

Cost/loss ratio: An idealised approach to decision-making from probability predictions which exploits the ratio between the cost of protection against adverse weather and the potential losses caused in the absence of protection – for a reliable ensemble the ratio provides the optimal decision threshold in terms of the probability of the weather event. Although idealised, has nevertheless been successfully used in some applications.

Discrimination: the ability of a prediction system to discriminate between occasions when an event will occur or will not occur. Discrimination is typically measured using the Relative Operating Characteristic (ROC) score.

Drift: The change of bias with prediction lead time.

Ensemble Mean: the mean (average) of a predicted variable or field over all the ensemble members.

Ensemble size: number of ensemble members.

Median: The value of a variable that divides a sample in half – the 50th percentile. There is equal probability that the value of a variable will be greater or smaller than the sample median. (taken from NOAA glossary).

Outlier: An observation which falls outside the range of values spanned by the ensemble members. It is normally considered for a correctly-spread ensemble with n members that a fraction of $2/(n+1)$ observations should be outliers; an ensemble with no outliers would likely be considered over-spread.

Percentile: the same as “quantile” but expressed as a percentage. For example, 30% of the values are less than the 30th percentile.

Predictability: The inherent degree to which the system is predictable. It corresponds to the maximum attainable skill with a perfect prediction system.

Quantile: the specific value of a variable that divides the distribution into two parts, those values greater than the quantile value and those values that are less. For instance, p percent of the values are less than the pth quantile. (taken from the NOAA glossary).

© Crown copyright 2024, Met Office

Spread: The standard deviation of ensemble members about the ensemble mean for a predicted variable. (The term spread is sometimes also used more subjectively to refer to whether an ensemble encompasses an event or scenario within its range of scenarios.)

Total Variance: The variance over all ensemble members and a series of start times. The total variance is an observable quantity and it should match the observed variability within statistical uncertainty.

Verification Metrics for Ensembles and Probability Predictions

Note that many of these have equivalent Skill scores which measure the skill of a forecast relative to a standard reference forecast, usually taken to be the climatological frequency of the event.

We do not provide the mathematical definitions of scores here but refer the reader to Jolliffe and Stephenson (2012) - https://doi.org/10.1002/9781119960003.

Brier (Skill) Score – measures quality of probability predictions for a specific event threshold. May be decomposed into Reliability, Resolution and Uncertainty (Skill) Scores. (See also https://glossary.ametsoc.org/wiki/Brier_score)

Continuous Ranked Probability (Skill) Score – Equivalent of RPS for continuous range of all possible thresholds. A good overall measure of ensemble skill.

Rank Histogram (sometimes formerly referred to as Talagrand Diagram) – illustrates the extent to which each ensemble member (or more precisely, the interval between each ensemble member) is equally likely; for a perfectly reliable ensemble each ranked interval, including the two outlier intervals, should be equally populated with observed values, giving a "flat" histogram.

Ranked Probability (Skill) Score – similar to a Brier score but for a range of thresholds rather than just one.

Relative (or Receiver) Operating Characteristic (ROC) - measures discrimination – ability to discriminate between an event occurring or not – for a specific event threshold. Based on Hit Rates and False Alarm Rates for a range of probability thresholds which are plotted on a graph. ROC area (area under the plotted ROC curve) provides a single figure to summarise the overall capability.

Relative Economic Value – based on ROC data combined with cost-loss information for a decision.

Reliability: The degree to which prediction probabilities accurately reflect the observed rate of occurrence. Reliability has a precise mathematical definition as part of the decomposition of the Brier Score. Reliability is a necessary but not sufficient requirement for good performance of ensemble or probability predictions – useful predictions require to have resolution in addition to reliability. Reliability of ensemble predictions is closely related to ensemble spread: an under-spread ensemble will display over-confidence where low prediction probabilities result in higher observed frequencies and high prediction probabilities result in lower observed frequencies (so that reliability diagram slope is less than ideal 45 degrees). An over-spread ensemble (rarer) would have a reliability diagram with >45 degree slope.

Reliability Diagram – visual representation of Brier Score and its decomposition – should be accompanied by a Sharpness diagram which plots the relative frequency of usage of different probability bins.

Resolution: In verification, a measure of how much the conditional probabilities differ from the climatic average. Resolution has a precise mathematical definition as part of the decomposition of the Brier Score. Higher resolution means that predictions have a large variability of the observed frequencies associated with different prediction probabilities around the climatological value. In the worst case, when the climatic probability is always prediction, the resolution is zero. (*Note: this is distinct from the term "model resolution" which typically refers to the grid-spacing of a numerical model and affects the degree of physical detail which can be explicitly resolved by the model.)*

© Crown copyright 2024, Met Office

Sharpness: Only applies to probability predictions. This is the extent to which an individual prediction differs from the overall average prediction. It can also be thought of as the degree to which probabilistic predictions approach categorical predictions of 0 or 1. (Taken from NOAA glossary).

Skill: Skill measures the relative performance of a prediction system (ensemble) compared to a reference prediction system. A skill score can be calculated for almost any verification metric (e.g. Brier Skill Score compares the Brier score of the predictions to the Brier Score of the reference.) The reference for probability predictions is typically the climatological frequency of the event, but for assessing an upgrade to an ensemble system it could be the score achieved by the previous version.

Spread-Error diagram – compares the growth in Ensemble Mean RMSE with the Ensemble Spread as a function of prediction lead-time.

© Crown copyright 2024, Met Office

Annex A: Acronyms and Abbreviations

BS – Brier Score
BSS – Brier Skill Score
C/L – Cost-loss
CRPS – Continuous Ranked probability Score
ENS – ECMWF Ensemble Prediction System
EPS – Ensemble Prediction System
IFS – ECMWF Integrated Forecast System
GEFS – NOAA Global Ensemble Forecast System
MOGREPS – Met Office Global and Regional Ensemble Prediction System
RMSE – Root Mean Square Error
REV – Relative Economic Value
RP – Random Parameters – a stochastic physics scheme
ROC – Relative (or Receiver) Operating Characteristic
RPS – Ranked Probability Score
SKEB - Stochastic Kinetic Energy Backscatter – a stochastic physics scheme
SPP – Stochastic Perturbed Parameters – a stochastic physics scheme
SPPT - Stochastic Perturbations of Physical Tendencies – a stochastic physics scheme

© Crown copyright 2024, Met Office